Learning Comprehensible Relational Features to Distinguish Subfossil Decapod Crustacean Dactyls

ILP2008

Mark Goadrich and Jeffrey Agnew
Departments of Computer Science and Geology
Centenary College of Louisiana

{mgoadric, jagnew}@centenary.edu

Abstract

Our research explores the application of Inductive Logic Programming to a new domain involving decapod crustacean claws. We find that we can distinguish dactyl shapes by automatically extracting relational features that describe their underlying spatial structure. We first use medial axis techniques to find the shock graph of each dactyl outline, which is then converted into a first-order logic representation capturing the connections, distances and angles between the nodes in this graph. We then use Aleph to find relational classification rules based on the shock graph representations. These relational rules provide a concise and human-understandable way to describe the morphological differences between closely related decapods, and can be seen as a first step to creating automatically learned quantitative taxonomic keys.

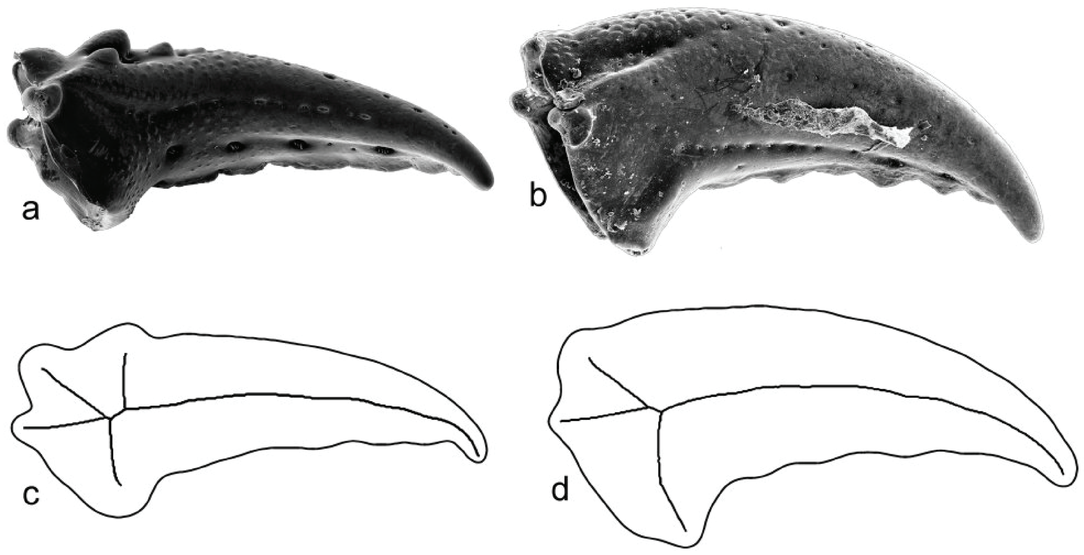


Figure 1: Scanning Electron Microscope images of Eriphia gonagra (a) and Menippe mercenaria (b) dactyls, and their corresponding shock graphs, (c) and (d).

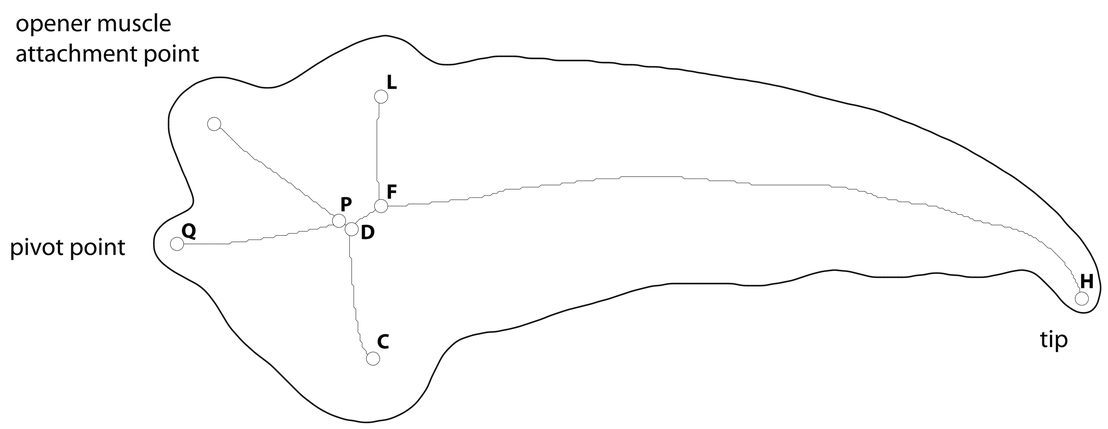
Task Description

Because decapod crustacean claws are potentially affected by numerous selective agents, they are excellent candidates for evolutionary studies of morphology. Despite being commonly found in shell-rich fossil assemblages, decapod dactyls (i.e., claw movable fingers) are usually ignored because of the assumption that they can be identified only to high taxonomic levels. Our research introduces a new method for distinguishing dactyl shapes by automatically extracting relational features that describe their underlying spatial structure, combining techniques from computer vision and inductive logic programming.

Our dataset for this study consists of 38 dactyl images, 12 belonging to *Eriphia gonagra* and 26 belonging to *Menippe mercenaria* [1]**.** We chose our head predicate to be the smaller class of *E. gonagra* dactyls. To create relational features for each dactyl and expose the underlying skeleton of the images, we chose to convert each image into a shock graph [2] using the flux skeleton implementation of ShapeMatcher [3]. A shock graph is created from a 2D image by first converting the image into an outline. This outline is then thinned along the normal vector according to the calculated flux at each point.

From this shock graph we create background knowledge shown in Table 1. Two predicates, `hasNode` and `hasEdge`, connect the nodes and edges with each example. We then calculate the `angle` between each adjacent edge, the `distance` between any two nodes in the graph, and include the predicates of >, <, >= and =< to compare these angles and distances. These basic predicates are enriched with definitions for acute, obtuse, interiorNode, and between to increase the expressive power of the learned rules.

Predicate Type	Predicate Name
Head	eriphia(+example).
Basic	hasNode(+example, -node). hasEdge(+example, -edge, -node, -node). angle(+edge, +edge, -float). distance(+node, +node, -float). (+float)>(+float). (+float)<(+float). (+float)=<(+float). (+float)>=(+float).
Acute	obtuse(+float). acute(+float).
Full	interiorNode(+node). between0and20(+float). . . . between280and300(+float).




```
eriphia(A):-
  hasEdge(A,B,C,D), hasEdge(A,E,D,F), hasEdge(A,G,F,H),
  distance(D,H,I), distance(F,C,J), J<I, hasEdge(A,K,F,L),
  distance(L,H,M), J<M, distance(L,C,N), J<N, hasEdge(A,O,P,Q),
  distance(Q,H,R), M<R, distance(L,Q,S), J<S.
```

Figure 2: Sample rule learned with Aleph on fold 0, with one possible match of the nodes when applied to an Eriphia gonagra example.

Results and Conclusions

We divided the data of 38 examples into five folds of roughly equal size, distributing the positive and negative examples separately to ensure a distribution in each subfold comparable to the complete dataset. We compare the results of using Aleph [4] with two baseline algorithms, one which classifies all examples as positive, and another which classifies all examples as negative. The true positive, false positive, true negative and false negative results across the five testsets are pooled to find the overall accuracy, precision, recall and F1 score for each algorithm.

Algorithm	Accuracy	Precision	Recall	F1 Score
Aleph	89.4	90.0	75.0	81.8
All Pos	31.6	31.6	100.0	48.0
All Neg	68.4	-	0.0	-

Figure 2 shows a sample rule learned with Aleph from fold 0 using only the basic background knowledge. This rule captures all of the positive *Eriphia gonagra* examples and none of the negative examples, in both the training set and testing set. It includes a sequence of connected nodes, C to D to F to H, where the distance between nodes C and F, called J, is less than other calculated distances in this rule. When this rule is applied to the positive examples, as seen in Figure 2, node C frequently corresponds to the closer muscle insertion point and H to the tip point.

We believe natural selection could be acting on the distances in this learned relationship between these areas of the shock graph. Because *Menippe* feeds almost exclusively on hard-shelled prey and *Eriphia* is more of an opportunistic generalist, *Menippe* should have claws with stronger biting forces than *Eriphia* [5]. Our learned rule discusses the length and angle of the closer muscle insertion point in relation to the tip. This relationship is directly related to the mechanical advantage of the claw, such that a shorter length in *E. gonagra* will result in weaker closing strength.

Our current dataset is quite small, with test folds having only 2 or 3 positive examples. We plan to further investigate this approach with a larger dataset consisting of 970 major and minor dactyls from nine xanthoid crab species. This dataset will allow us to evaluate whether this method can be used to distinguish dactyls of several closely related species. Also, because many of the dactyls of these species change shape with growth, we can quantify those allometric transformations and identify dactyl sizes where species level differences emerge.

References

1. J. G. Agnew. *Dactyls Reveal Evolutionary Patterns in Decapod Crustaceans*. PhD thesis, Louisiana State University - Baton Rouge, 2008.

2. P. Dimitrov, C. Phillips, and K. Siddiqi. Robust and efficient skeletal graphs. In *Computer Vision and Pattern Recognition*, 2000.

3. D. Macrini. *Indexing and matching for view-based 3-D object recognition using shock graphs*. Master's thesis, University of Toronto, July 2003.

4. A. Srinivasan. The Aleph Manual Version 5. http://web.comlab.ox.ac.uk/oucl/research/areas/ machlearn/Aleph/, 2003.

5. A. B. Williams. *Shrimps, lobsters, and crabs of the Atlantic coast of the eatern United States, Maine to Florida.* Smithsonian Institution Press, Washington, D.C., 1984.

We would like to thank the authors of the software packages Aleph, SHAPE and ShapeMatcher for the availability of their code, and Centenary College for providing travel funds.